

УДК 811.112'37

## **СОВРЕМЕННЫЕ ИНСТРУМЕНТЫ КОНТЕНТ-АНАЛИЗА МЕДИЙНЫХ ТЕКСТОВ**

**О. Г. Чупрына, Д. О. Короткевич**

*ГАОУ ВО МГПУ*

*Москва, Российская Федерация*

*[chuprinaog@mgpu.ru](mailto:chuprinaog@mgpu.ru)*

Статья посвящена применению программы Wmatrix как современного инструмента для компьютерного контент-анализа медиатекстов. На материале публикаций американских газет выявляются когнитивные контексты, лексико-семантические поля и типовые аргументы, передающие ключевые смысловые установки.

*Ключевые слова:* медиалингвистика, контент-анализ, когнитивный контекст, лексико-семантическое поле, медиарепрезентация, типовой аргумент, *Wmatrix*.

Наиболее заметными тенденциями в современной лингвистике стали изучение цифровых медийных текстов и использование цифровых инструментов обработки текста. Первая тенденция базируется на потребности разработки методов изучения способов воздействия на общественное мнение с помощью языковых средств в каналах массовой коммуникации. Вторая сформировалась в результате соединения компьютерных технологий и методов интерпретации текста. Автоматизация процесса сбора и обработки лингвистических данных позволяет не только ускорить исследовательскую работу, но и повысить её объективность и воспроизводимость. Исследователями был апробирован комплексный анализ текстов статей, посвященных Бразилии, в американских газетах *The New York Times* и *The New York Post* за 2019 и 2023 год [4]. Он включает использование программы *Wmatrix* – одного из наиболее универсальных и функциональных инструментов для компьютерного контент-анализа. Цель статьи состоит в представлении результатов применения компьютерного контент-анализа газетных статей на основе программы *Wmatrix*.

*Wmatrix* – это программный комплекс, разработанный на базе Ланкастерского университета (UCREL), который позволяет проводить автоматическую разметку текстов на уровне части речи, семантической категории (USAS), а также выполнять лексико-семантический и контекстный анализ. Программа включает в себя встроенные функции частотного анализа, выделения ключевых слов (keyness), сравнения корпусов и построения графиков по результатам обработки. Использование подобных цифровых платформ особенно актуально при работе с большими объемами аутентичных текстов, когда традиционные методы анализа оказываются недостаточно эффективными. Программное обеспечение *Wmatrix* открывает возможности для систематизации и интерпретации семантических и прагматических характеристик текста, что особенно ценно для медиалингвистики. Такой подход позволяет объединить количественный и качественный анализ, обеспечивая комплексное понимание лингвистических механизмов, задействованных в медиарепрезентации.



Рис. 1. Лексико-семантические поля (ЛСП) матрицы Бразилия, выделенные при помощи программы *Wmatrix* по материалам газеты *The New York Times* (2023)

Основное преимущество *Wmatrix* – это возможность анализа текстов на семантическом уровне: каждый лексемный элемент автоматически соотносится с определённой семантической категорией. Опираясь на определение Ю. Н. Караулова, согласно которому лексико-семантическое поле представляет собой группу слов, обладающих семантическим сходством и являющихся различными способами выражения одного и того же языкового понятия [1], представляется обоснованным называть выявляемые программой *Wmatrix* группы лексем именно лексико-семантическими полями (ЛСП) (рис. 1). Это позволяет выделить лексико-семантические поля и составляющие их лексические единицы, характерные для материалов о Бразилии в американских газетах *The New York Times* и *The New York Post*, а также выявить, обращение к каким когнитивным контекстам реализуется в медиа репрезентации Бразилии в разные периоды (2019 и 2023 гг.) (рис. 2) [2].

Word	Semtag	Frequency	Relative Frequency	Concordance	Summary information:																																				
fraud	G2.1-	32	0.04	Concordance	<b>Number of types shown: 67</b> <b>Total frequency of types shown: 228 (0.31%)</b> <b>Total frequency overall: 74229</b>  <b>Number of items shown with a given frequency:</b> <table border="1"> <thead> <tr> <th>Frequency</th> <th>Types</th> <th>Tokens</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>39 (58.21%)</td> <td>39 (17.11%)</td> </tr> <tr> <td>2</td> <td>12 (17.91%)</td> <td>24 (10.53%)</td> </tr> <tr> <td>3</td> <td>3 (4.48%)</td> <td>9 (3.95%)</td> </tr> <tr> <td>4</td> <td>3 (4.48%)</td> <td>12 (5.26%)</td> </tr> <tr> <td>5</td> <td>(0.00%)</td> <td>(0.00%)</td> </tr> <tr> <td>6</td> <td>4 (5.97%)</td> <td>24 (10.53%)</td> </tr> <tr> <td>7</td> <td>(0.00%)</td> <td>(0.00%)</td> </tr> <tr> <td>8</td> <td>1 (1.49%)</td> <td>8 (3.51%)</td> </tr> <tr> <td>9</td> <td>(0.00%)</td> <td>(0.00%)</td> </tr> <tr> <td>10</td> <td>(0.00%)</td> <td>(0.00%)</td> </tr> <tr> <td>&gt; 10</td> <td>5 (7.46%)</td> <td>112 (49.12%)</td> </tr> </tbody> </table>	Frequency	Types	Tokens	1	39 (58.21%)	39 (17.11%)	2	12 (17.91%)	24 (10.53%)	3	3 (4.48%)	9 (3.95%)	4	3 (4.48%)	12 (5.26%)	5	(0.00%)	(0.00%)	6	4 (5.97%)	24 (10.53%)	7	(0.00%)	(0.00%)	8	1 (1.49%)	8 (3.51%)	9	(0.00%)	(0.00%)	10	(0.00%)	(0.00%)	> 10	5 (7.46%)	112 (49.12%)
Frequency	Types	Tokens																																							
1	39 (58.21%)	39 (17.11%)																																							
2	12 (17.91%)	24 (10.53%)																																							
3	3 (4.48%)	9 (3.95%)																																							
4	3 (4.48%)	12 (5.26%)																																							
5	(0.00%)	(0.00%)																																							
6	4 (5.97%)	24 (10.53%)																																							
7	(0.00%)	(0.00%)																																							
8	1 (1.49%)	8 (3.51%)																																							
9	(0.00%)	(0.00%)																																							
10	(0.00%)	(0.00%)																																							
> 10	5 (7.46%)	112 (49.12%)																																							
criminal	G2.1-	26	0.04	Concordance																																					
stolen	G2.1-	25	0.03	Concordance																																					
crimes	G2.1-	17	0.02	Concordance																																					
crime	G2.1-	12	0.02	Concordance																																					
plot	G2.1-	8	0.01	Concordance																																					
conspiracy	G2.1-	6	0.01	Concordance																																					
illegal	G2.1-	6	0.01	Concordance																																					
felony	G2.1-	6	0.01	Concordance																																					
illegally	G2.1-	6	0.01	Concordance																																					
broke_into	G2.1-	4	0.01	Concordance																																					
terrorism	G2.1-	4	0.01	Concordance																																					
violation	G2.1-	4	0.01	Concordance																																					
wrongdoing	G2.1-	3	0.00	Concordance																																					
stole	G2.1-	3	0.00	Concordance																																					
steal	G2.1-	3	0.00	Concordance																																					
stealing	G2.1-	2	0.00	Concordance																																					
injustice	G2.1-	2	0.00	Concordance																																					
robbed	G2.1-	2	0.00	Concordance																																					
terrorists	G2.1-	2	0.00	Concordance																																					
break_into	G2.1-	2	0.00	Concordance																																					
fraudulent	G2.1-	2	0.00	Concordance																																					
breaking_the_law	G2.1-	2	0.00	Concordance																																					
racketeering	G2.1-	2	0.00	Concordance																																					
terrorist	G2.1-	2	0.00	Concordance																																					
vandalism	G2.1-	2	0.00	Concordance																																					
robberies	G2.1-	2	0.00	Concordance																																					
outlaws	G2.1-	2	0.00	Concordance																																					
against_the_law	G2.1-	1	0.00	Concordance																																					
criminality	G2.1-	1	0.00	Concordance																																					
conspiring	G2.1-	1	0.00	Concordance																																					
looting	G2.1-	1	0.00	Concordance																																					
subversive	G2.1-	1	0.00	Concordance																																					
kidnappings	G2.1-	1	0.00	Concordance																																					
offenses	G2.1-	1	0.00	Concordance																																					
suspects	G2.1-	1	0.00	Concordance																																					
rape	G2.1-	1	0.00	Concordance																																					
is_up_for	G2.1-	1	0.00	Concordance																																					
theft	G2.1-	1	0.00	Concordance																																					
treason	G2.1-	1	0.00	Concordance																																					
guilty	G2.1-	1	0.00	Concordance																																					

Рис. 2. Единицы ЛСП Crime, реализуемые в газете *The New York Times* (2023)

В исследуемых газетных публикациях за оба периода было установлено наличие ряда ведущих когнитивных контекстов, каждый из которых коррелирует с определёнными лексико-семантическими полями.

1) Когнитивный контекст *Social Dynamics* реализуется через такие лексико-семантические поля, как *Personal Names*, *Belonging to a Group*, *Dead* и *Helping*.

2) Когнитивный контекст *Geography* представлен активным функционированием лексико-семантических полей *Geographical Names*, *Places* и *Geographical Terms*.

3) Когнитивный контекст *State Organisation* связан с такими лексико-семантическими полями, как *Government*, *In Power* и *Politics*.

4) Когнитивный контекст *Education* представлен в материалах через лексико-семантическое поле *Education*.

5) Когнитивный контекст *Ecology* формируется за счёт лексико-семантических полей *Temperature: Hot/on fire* и *Green Issues*.

6) Когнитивный контекст *Justice* проявляется через единицы лексико-семантических полей *Law and Order*, *Warfare: Defence and the Army/Weapons*, *Violent/Angry*, *Crime* и *Hindering*.

В сочетании с анализом конкорданса (контекстуального окружения ключевых лексем) использование программы *Wmatrix* позволяет глубже проникнуть в типовые структуры медиарепрезентации и выявить механизмы формирования устойчивых смыслов (рис. 3). Одним из таких механизмов является типовая аргументация, в основе которой лежат повторяющиеся абстрактные структуры – типовые аргументы. Каждый типовой аргумент включает три элемента: данные, правило и вывод; при этом данные могут быть выражены имплицитно, а правило может быть опущено [3]. Повторяемость структуры «данные – правило – вывод» способствует закреплению в сознании аудитории определенных интерпретаций и смыслов.

17 occurrences.		Extend context
s from horseback and filming their	crimes as they committed them . We always	1 More   Full
sts and will pay for the committed	crimes . Rita Lee , Brazils Queen of Rock	2 More   Full
nvestigating potential Russian war	crimes in Ukraine . But when he arrived at	3 More   Full
circles . He was charged with hate	crimes , but the case was later dismissed	4 More   Full
of spreading Nazi ideology . Both	crimes can lead to yearslong prison terms	5 More   Full
e could be charged with committing	crimes against democratic institutions or	6 More   Full
ermath would pay for the committed	crimes . Mr. Lula signed an emergency decr	7 More   Full
. Bolsonaro should be charged with	crimes against humanity for his handling o	8 More   Full
dictatorship-era torture and other	crimes and a review of the law that provid	9 More   Full
ad committed politically motivated	crimes in the previous two decades , cover	10 More   Full
have ever been punished for their	crimes . Thats why Brazilians can not watc	11 More   Full
sand people have been convicted of	crimes against humanity in our neighboring	12 More   Full
rs of the dictatorship hint at the	crimes of the other side the leftist guerr	13 More   Full
military never apologized for its	crimes . On the contrary , it still celebr	14 More   Full
uch posts could effectively incite	crimes , and thus the court should investi	15 More   Full
ndowski said . Those who committed	crimes against the public have to stay out	16 More   Full
ovement co-organizer , of inciting	crimes with his call for new occupations ,	17 More   Full

Рис. 3. Пример конкорданса ядра ЛСП Crime в газете The New York Times (2023)

Примером реализации подобной структуры служит аргумент «Болсонару — ультраправый политик», выявленный во всех исследуемых газетах. В публикациях The New York Times и The New York Post политическая принадлежность Жаира Болсонару маркируется явно оценочной лексикой с негативной коннотацией:

- ‘*Mr. Bolsonaro, a far-right populist and climate change skeptic*’ [NYT 2019] («Г-н Болсонару, ультраправый популист и скептик в вопросах изменения климата»);
- ‘*Controversial far-right politician Jair Bolsonaro wins Brazil’s presidential election*’ [NYP 2019] («Противоречивый ультраправый политик Жаир Болсонару побеждает на президентских выборах в Бразилии»).

Данные фрагменты демонстрируют использование единиц ЛСП *Politics (far-right populist, politician)* в сочетании с эмотивно окрашенными номинациями (*controversial*), которые одновременно информируют и формируют у читателя определённое оценочное восприятие фигуры Ж. Болсонару. Таким образом, структура аргумента строится следующим образом: факт (указание на политическую ориентацию) → правило (связь ультраправых взглядов с определённым типом политического поведения) → вывод (формирование критической оценки личности политика).

Таким образом, *Wmatrix* может рассматриваться как универсальный инструмент для исследователей, работающих с медиатекстами. Он позволяет не только классифицировать языковые единицы, но и интерпретировать их в контексте когнитивных структур, идеологических установок и жанровой специфики. В заключение можно отметить, что использование программных платформ контент-анализа не заменяет интерпретацию, но создаёт прочную количественную и категориальную основу для лингвистических выводов, особенно в исследовании динамики смыслов в медиадискурсе.

#### БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Караулов, Ю. Н. Структура лексико-семантического поля / Ю. Н. Караулов // Филологические науки. – 1972. – № 1. – С. 57.
2. Короткевич, Д. О. Лексико-семантическое поле ‘Education’ как компонент когнитивной матрицы «Бразилия» (на материале американских СМИ) / Д. О. Короткевич // Вестник филологических наук. – 2023. – Т. 3, № 7. – С. 179-186.
3. Спасская А. С. Типовой аргумент "Йорг Хайдер - правый популист" в системе дискурсивного конструирования немецкими СМИ идентичности австрийцев (на материале публикаций газеты "Der Tagesspiegel" в феврале 2000 г.) // Филологические науки. Вопросы теории и практики. 2023. №5. С. 1515-1524.
4. Чупрына, О. Г. Когнитивно-матричный анализ репрезентации концепта Бразилия в дискурсе СМИ (на материале газет The New York Times и The New York Post) / О. Г. Чупрына, Д. О. Короткевич // Иностранные языки в высшей школе. – 2022. – № 3(62). – С. 54-62.

**MODERN TOOLS OF CONTENT ANALYSIS OF MEDIA TEXTS (ON THE  
EXAMPLE OF WMATRIX PROGRAM)**

**O. G. Chupryna, D. O. Korotkevich**

*Moscow City University*

*Moscow, Russia*

*chuprinaog@mgpu.ru*

The article explores the use of Wmatrix as a modern tool for computer-assisted content analysis of media texts. Drawing on material from American newspaper publications, the study identifies cognitive contexts, lexical-semantic fields, and typical arguments that convey key ideological and semantic frameworks.

*Keywords:* media linguistics, content analysis, cognitive context, lexical-semantic field, media representation, typical argument, Wmatrix.